Center for Surveillance, Epidemiology, and Laboratory Services

Data Quality

Roseanne English, BS

Analytic Data Management Team Lead

NSSP Grantee Meeting

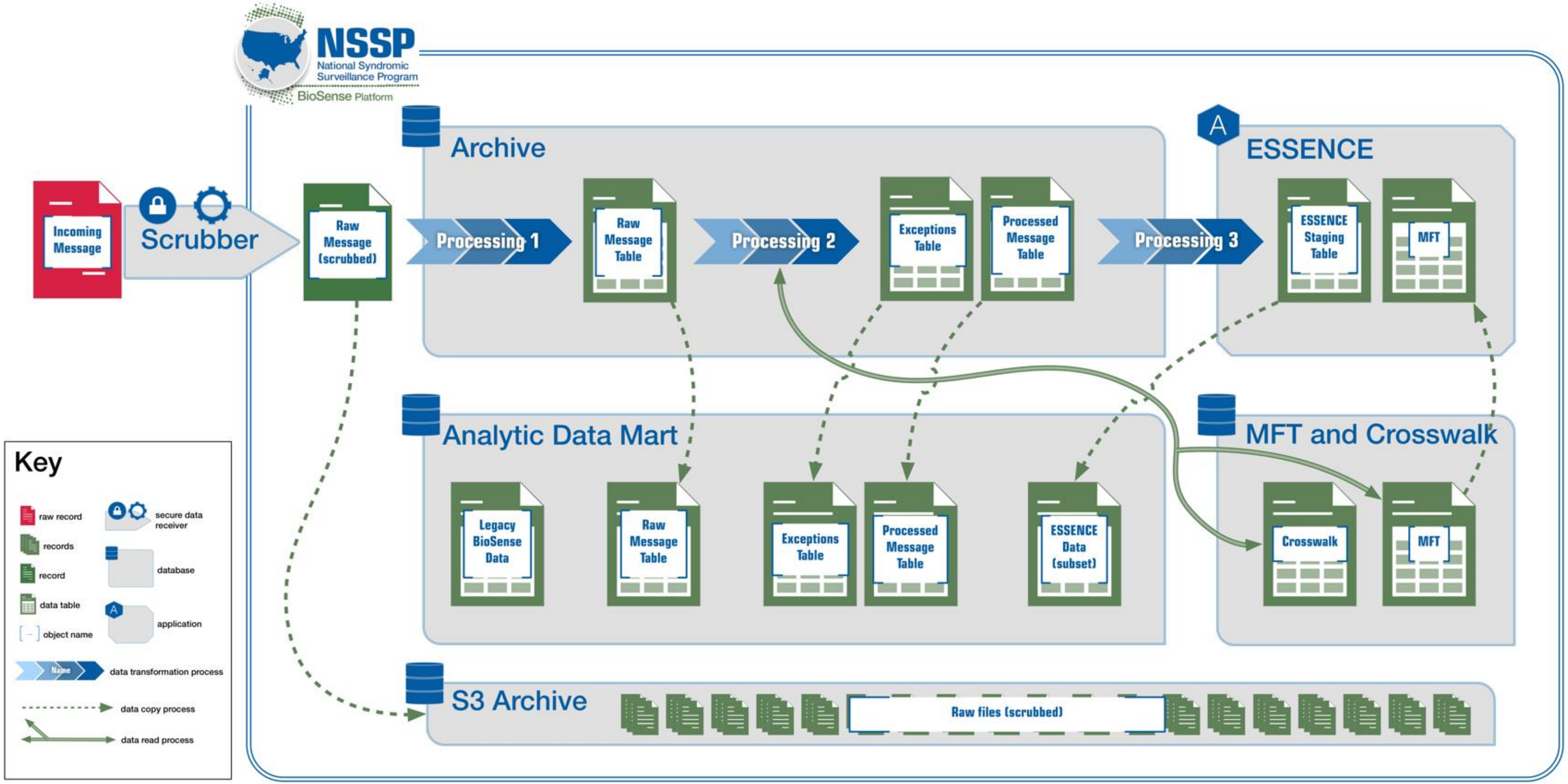
February 2017

Overview

- High-level Review of Data Flow
- Foundational Data Quality (DQ)
- Deeper Dive DQ Review of Data Content
- Feedback from the Community
- Next Steps

Overview – Data Flow

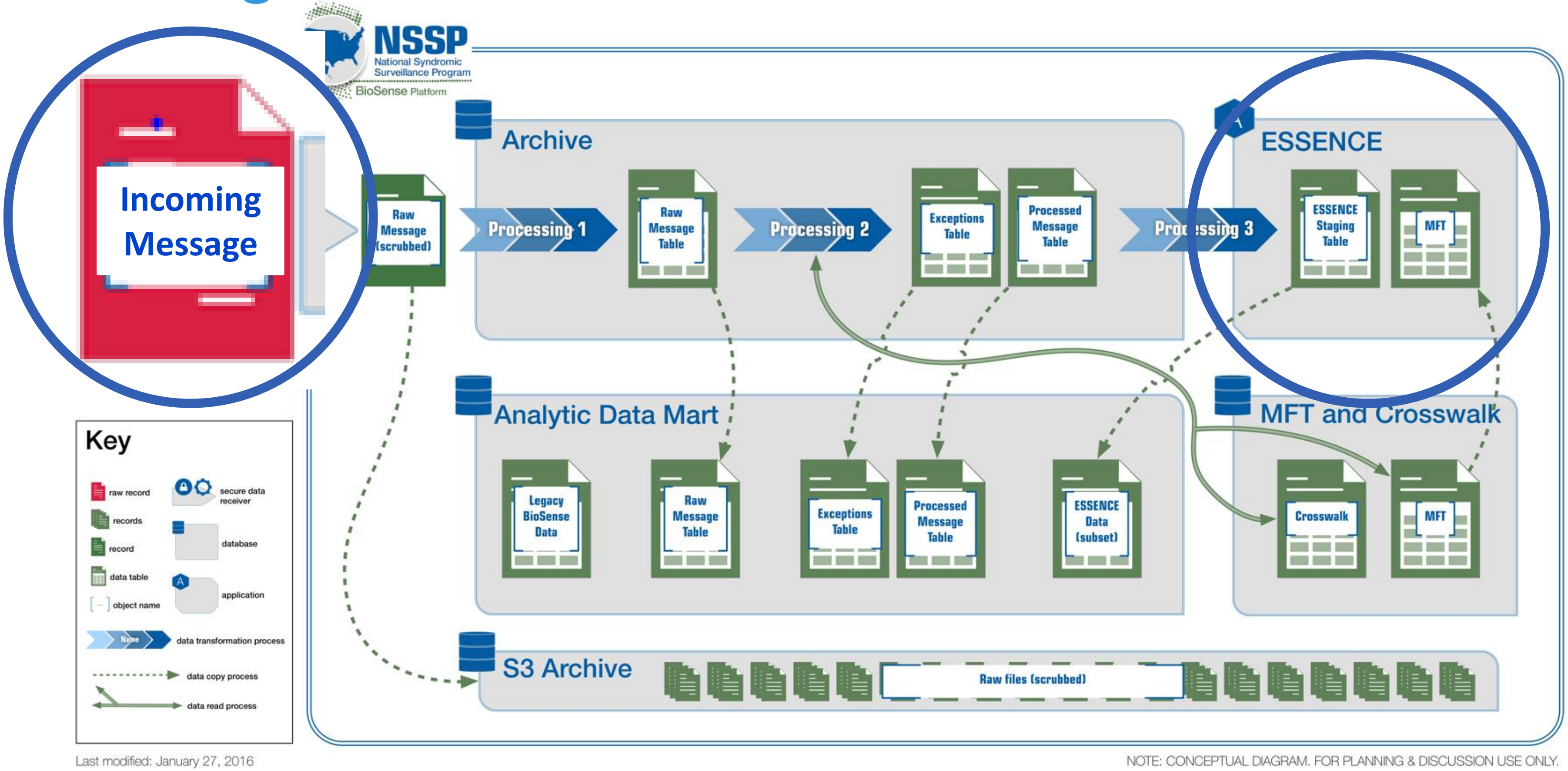
- High-level Review of Data Flow
- Foundational Data Quality (DQ)
- Deeper Dive DQ Review of Data Content
- Feedback from the Community
- Next Steps



Last modified: January 27, 2016

NOTE: CONCEPTUAL DIAGRAM. FOR PLANNING & DISCUSSION USE ONLY.

Incoming Data from Jurisdictions

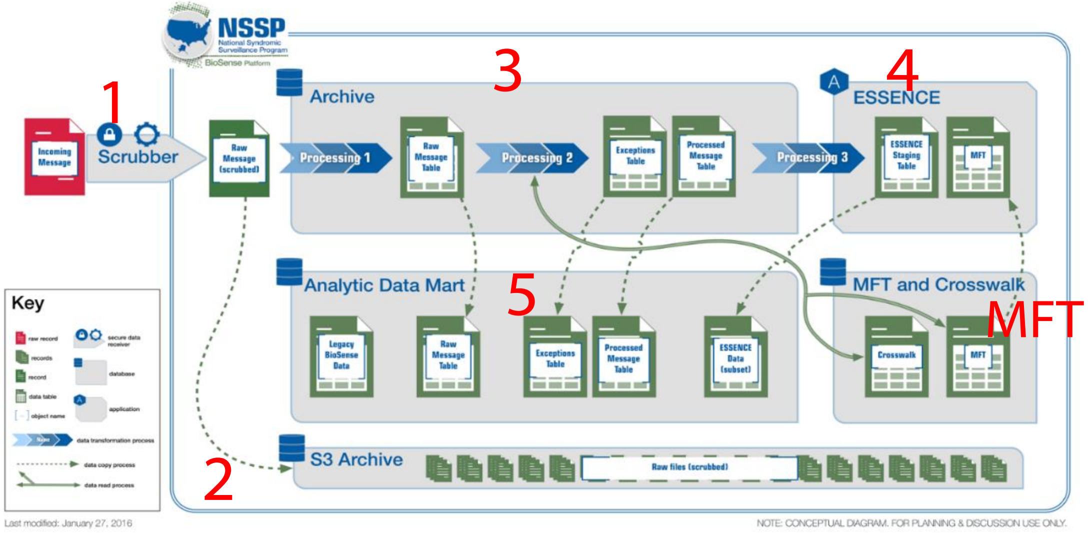


Last modified: January 27, 2016

NOTE: CONCEPTUAL DIAGRAM. FOR PLANNING & DISCUSSION USE ONLY.

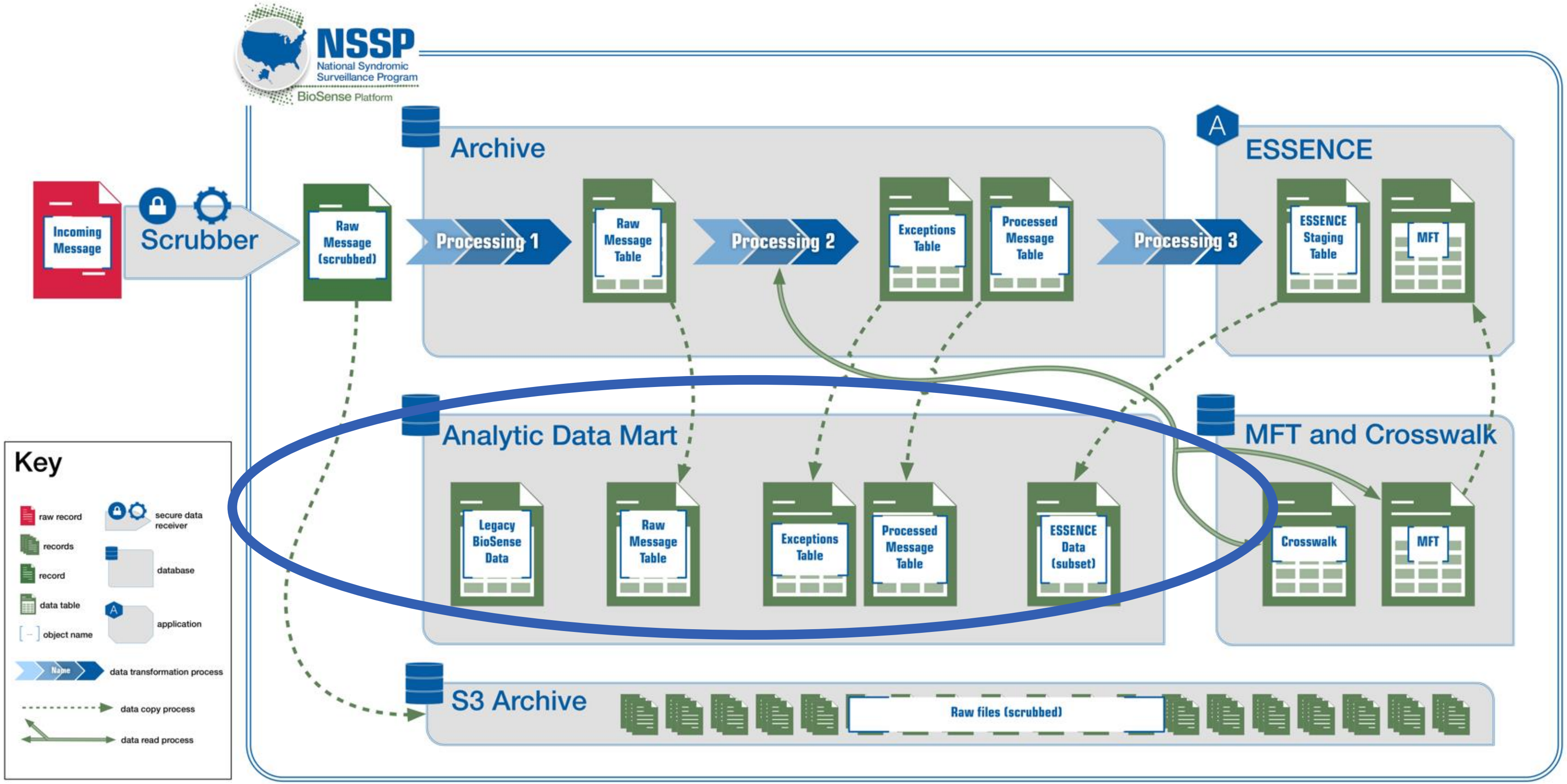
NSSP Process Components

1. “Scrub” targeted HL7 message components and incoming data to remove PII
2. Store archival copies of incoming data files
3. Ingest data into the *BioSense Platform Archive Database*
4. Ingest data into the *ESSENCE application*
5. Replicate data to an *Analytic Data Mart* to support complex analyses



Maintain Master Facility Table data and Crosswalk information (ongoing)

The “Bread and Butter” Used in DQ



NSSP
National Syndromic Surveillance Program
BioSense Platform

Incoming Message

Scrubber

Raw Message (scrubbed)

Archive

Processing 1

Raw Message Table

Processing 2

Exceptions Table

Processed Message Table

Processing 3

ESSENCE

ESSENCE Staging Table

MFT

Analytic Data Mart

Legacy BioSense Data

Raw Message Table

Exceptions Table

Processed Message Table

ESSENCE Data (subset)

MFT and Crosswalk

Crosswalk

MFT

S3 Archive

Raw files (scrubbed)

Key

raw record

secure data receiver

records

record

database

data table

application

[--] object name

Name

data transformation process

data copy process

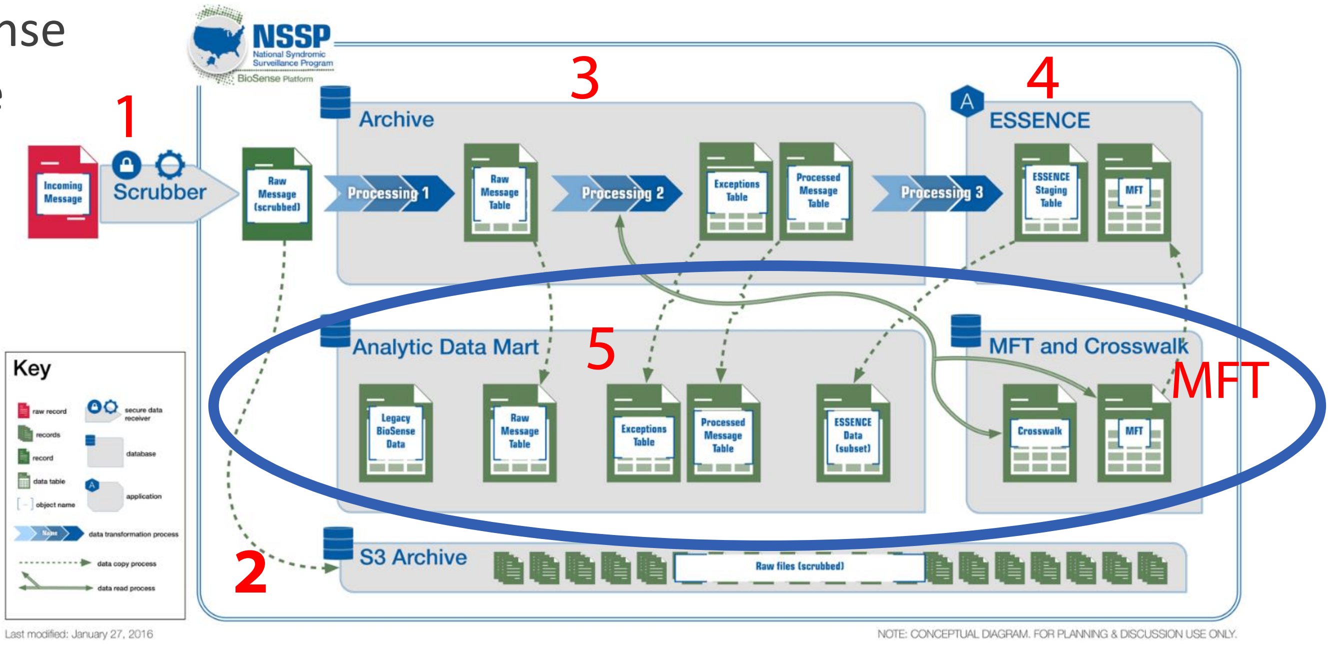
data read process

Last modified: January 27, 2016

NOTE: CONCEPTUAL DIAGRAM. FOR PLANNING & DISCUSSION USE ONLY.

Analytic Data Mart

- Replication of “bread and butter” BioSense ARCHIVE database
- Tools to access Analytic Data Mart for DQ and other complex analysis
  - ADMINER
  - R Studio Pro
  - SAS Studio (assessing)

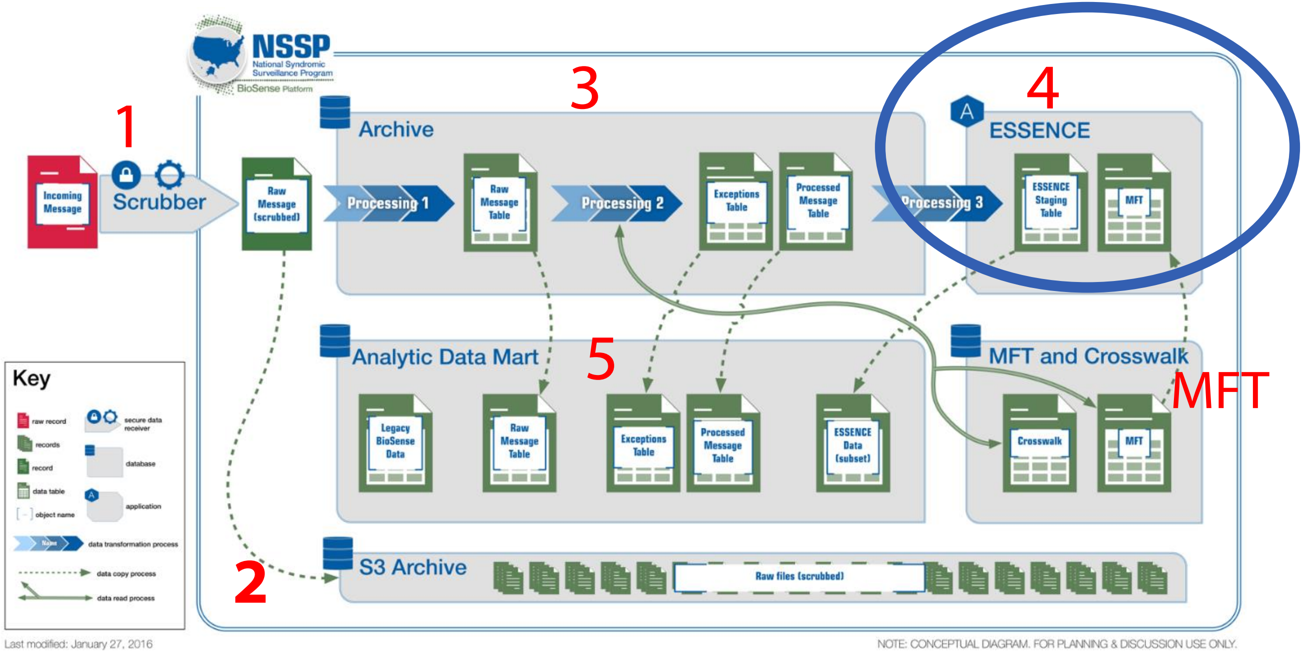


Last modified: January 27, 2016

NOTE: CONCEPTUAL DIAGRAM. FOR PLANNING & DISCUSSION USE ONLY.

Interactive ESSENCE Application for Surveillance

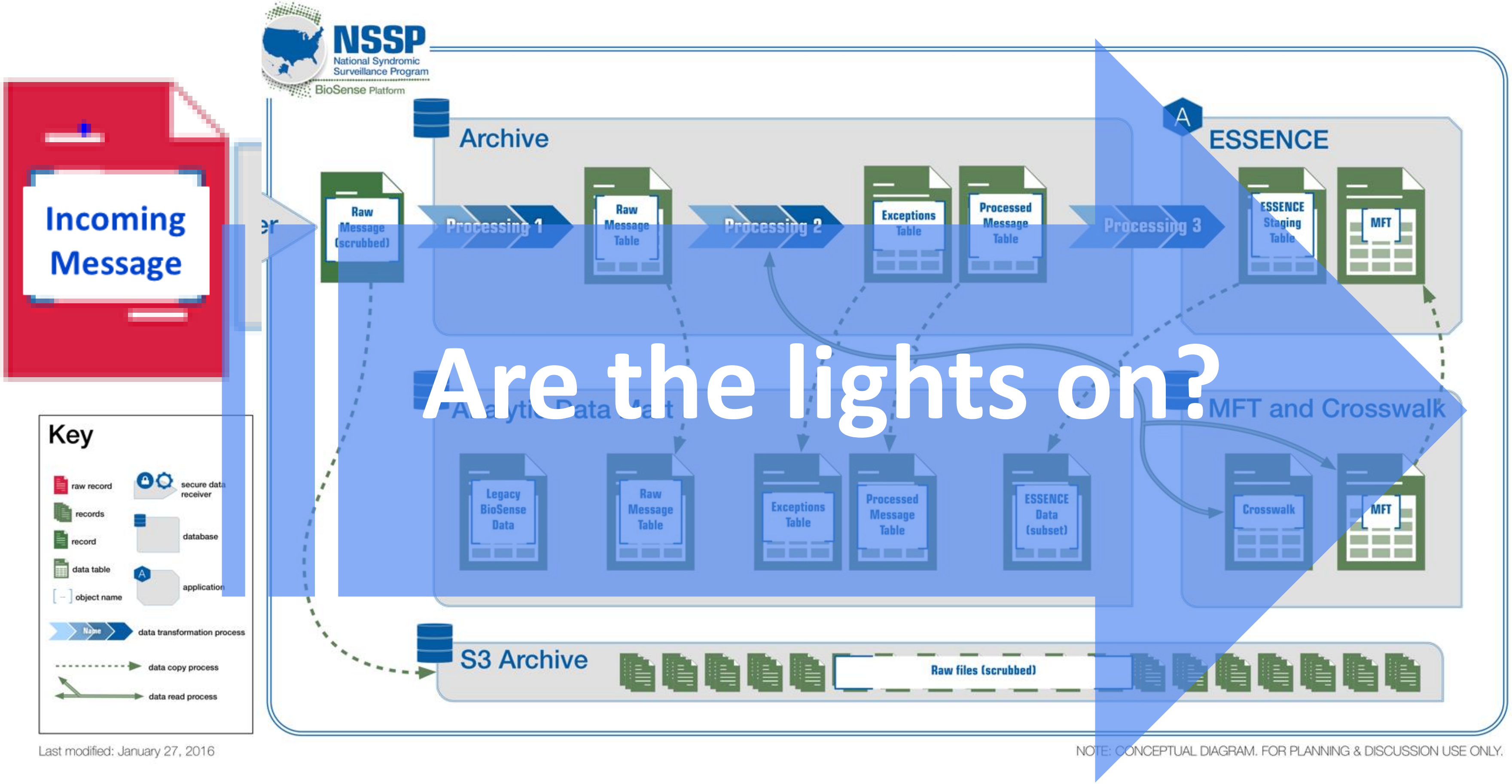
- A subset of fields from the BioSense Platform Archive
- Additional business rules are applied as data are formatted and transformed for use with ESSENCE



Overview - DQ

- High-level Review of Data Flow
- Foundational Data Quality (DQ)
- Deeper Dive DQ Review of Data Content
- Feedback from the Community
- Next Steps

Data Flow/Volume Checks



Incoming Message

Are the lights on?

Incoming Data

Daily report process checking incoming HL-7/ASCII feeds by site and feed name:

- Date last received
- Daily volume received
- Deviation in average records/visits received (in progress)

Incoming Data to ARCHIVE
Data Flow Checks by Site

Daily Report Process checks RAW, Processed, Exceptions data

- Maximum Create Date
- Lag time between Maximum Create Date and Date of DQ Report
- Percent of records that . . .
  - Filtered (RAW)
  - Excepted (Exceptions)
  - Successfully processed (Processed)

Incoming Data to ARCHIVE Data Flow Checks by Site

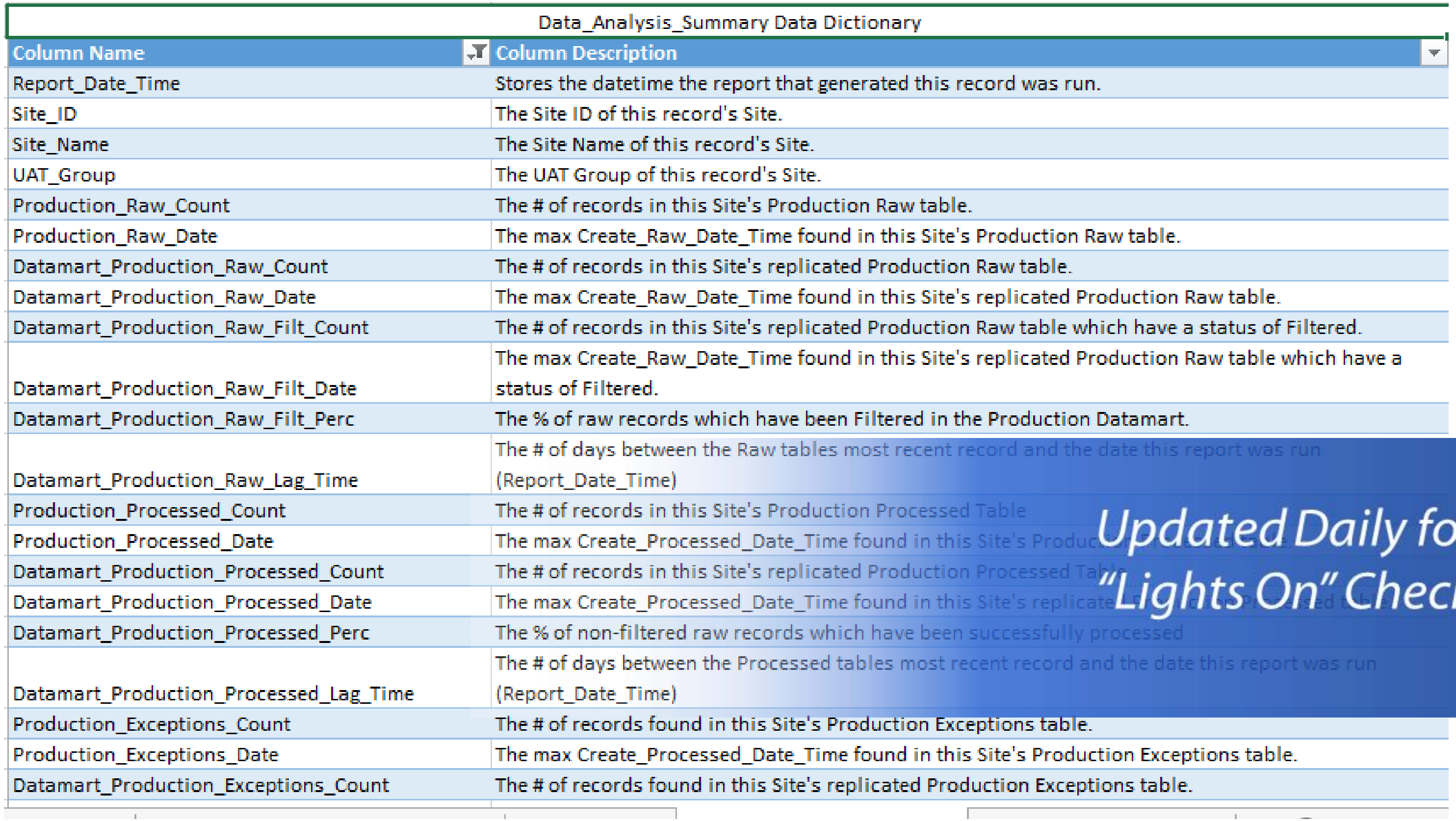
Automatic Alerts

- Volume discrepancies from one “data stop” to the next
- Processing lag time more than 24 hours from one “data stop” to the next
- High percent of
  - Filtered
  - Excepted

Action

- Generate auto-emails to internal team
- Determine root of the problem
- Alert and engage site as appropriate

Excerpt of Reporting Database



Data_Analysis_Summary Data Dictionary

Column Name	Column Description
Report_Date_Time	Stores the datetime the report that generated this record was run.
Site_ID	The Site ID of this record's Site.
Site_Name	The Site Name of this record's Site.
UAT_Group	The UAT Group of this record's Site.
Production_Raw_Count	The # of records in this Site's Production Raw table.
Production_Raw_Date	The max Create_Raw_Date_Time found in this Site's Production Raw table.
Datamart_Production_Raw_Count	The # of records in this Site's replicated Production Raw table.
Datamart_Production_Raw_Date	The max Create_Raw_Date_Time found in this Site's replicated Production Raw table.
Datamart_Production_Raw_Filt_Count	The # of records in this Site's replicated Production Raw table which have a status of Filtered.
Datamart_Production_Raw_Filt_Date	The max Create_Raw_Date_Time found in this Site's replicated Production Raw table which have a status of Filtered.
Datamart_Production_Raw_Filt_Perc	The % of raw records which have been Filtered in the Production Datamart.
Datamart_Production_Raw_Lag_Time	The # of days between the Raw tables most recent record and the date this report was run (Report_Date_Time)
Production_Processed_Count	The # of records in this Site's Production Processed Table
Production_Processed_Date	The max Create_Processed_Date_Time found in this Site's Production Processed table.
Datamart_Production_Processed_Count	The # of records in this Site's replicated Production Processed Table
Datamart_Production_Processed_Date	The max Create_Processed_Date_Time found in this Site's replicated Production Processed table.
Datamart_Production_Processed_Perc	The % of non-filtered raw records which have been successfully processed
Datamart_Production_Processed_Lag_Time	The # of days between the Processed tables most recent record and the date this report was run (Report_Date_Time)
Production_Exceptions_Count	The # of records found in this Site's Production Exceptions table.
Production_Exceptions_Date	The max Create_Processed_Date_Time found in this Site's Production Exceptions table.
Datamart_Production_Exceptions_Count	The # of records found in this Site's replicated Production Exceptions table.

Updated Daily for "Lights On" Checks

ARCHIVE to ESSENCE
Data Flow Checks by Site

Daily Report Process checking ESSENCE Ingestion

- Maximum Create Date
- Lag time between Maximum Create Date in ARCHIVE vs. ESSENCE
- Total count of records (ER_Import_Staging; ER_Base)
- Volume discrepancies between ARCHIVE and ESSENCE

ARCHIVE to ESSENCE
Data Flow Checks by Site Contd.

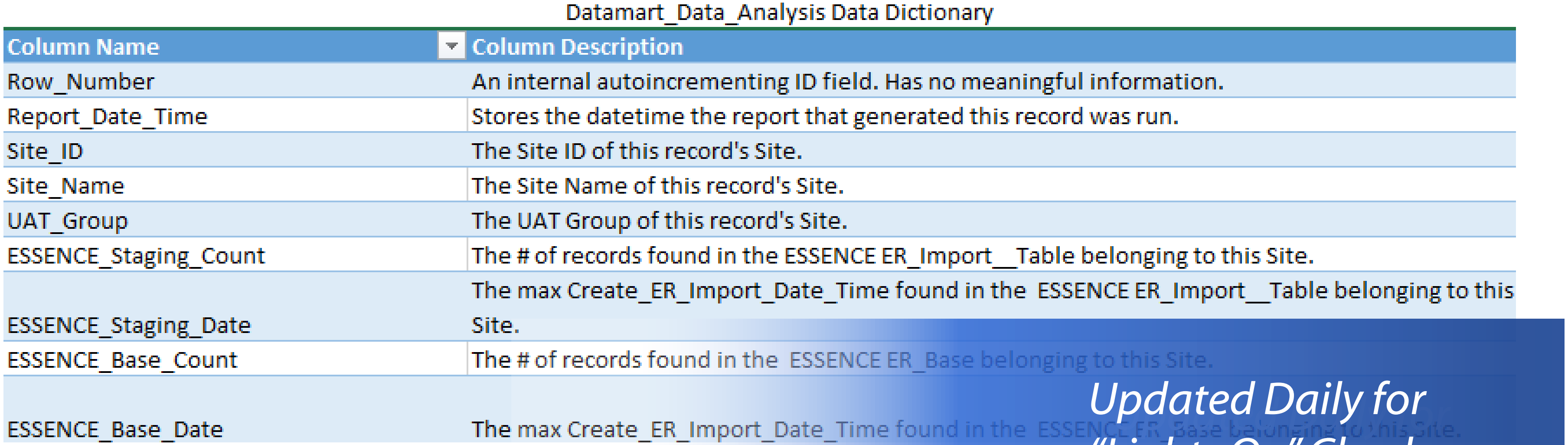
Automatic Alerts

- Volume discrepancies from ARCHIVE to ESSENCE
- Processing Lag time over 24 hours

Action

- Automatically alert internal team via email
- Determine root of the problem
- Alert and engage ESSENCE colleagues as appropriate

Excerpt of Reporting Database – Data Dictionary

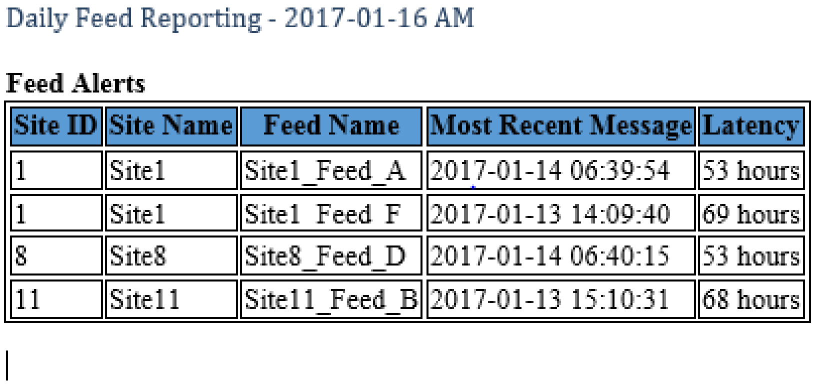


Datamart_Data_Analysis Data Dictionary

Column Name	Column Description
Row_Number	An internal autoincrementing ID field. Has no meaningful information.
Report_Date_Time	Stores the datetime the report that generated this record was run.
Site_ID	The Site ID of this record's Site.
Site_Name	The Site Name of this record's Site.
UAT_Group	The UAT Group of this record's Site.
ESSENCE_Staging_Count	The # of records found in the ESSENCE ER_Import__Table belonging to this Site.
ESSENCE_Staging_Date	The max Create_ER_Import_Date_Time found in the ESSENCE ER_Import__Table belonging to this Site.
ESSENCE_Base_Count	The # of records found in the ESSENCE ER_Base belonging to this Site.
ESSENCE_Base_Date	The max Create_ER_Import_Date_Time found in the ESSENCE ER_Base belonging to this Site.

Updated Daily for “Lights On” Checks

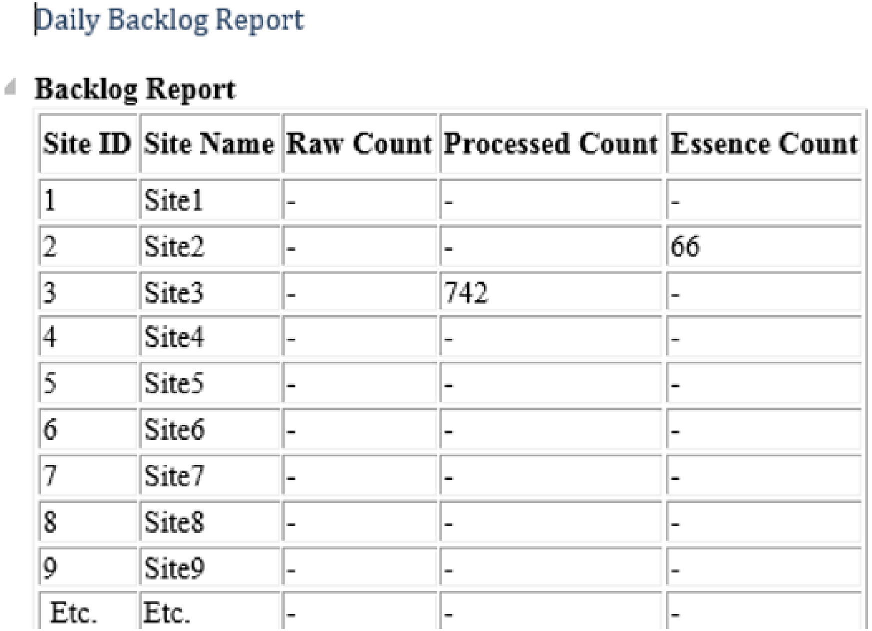
Examples of Alerts



Daily Feed Reporting - 2017-01-16 AM

Feed Alerts

Site ID	Site Name	Feed Name	Most Recent Message	Latency
1	Site1	Site1_Feed_A	2017-01-14 06:39:54	53 hours
1	Site1	Site1_Feed_F	2017-01-13 14:09:40	69 hours
8	Site8	Site8_Feed_D	2017-01-14 06:40:15	53 hours
11	Site11	Site11_Feed_B	2017-01-13 15:10:31	68 hours



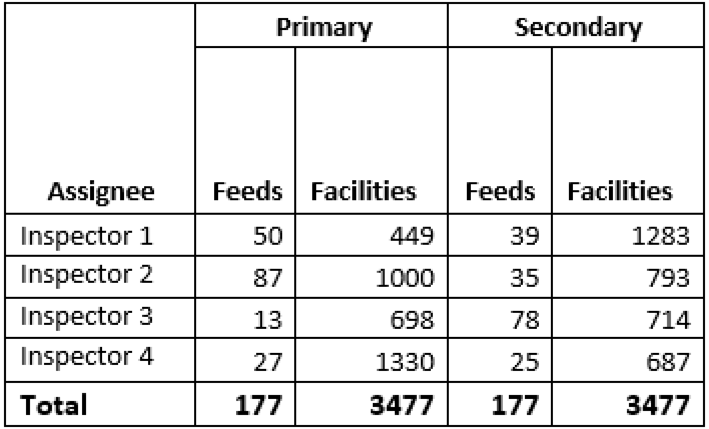
Daily Backlog Report

Backlog Report

Site ID	Site Name	Raw Count	Processed Count	Essence Count
1	Site1	-	-	-
2	Site2	-	-	66
3	Site3	-	742	-
4	Site4	-	-	-
5	Site5	-	-	-
6	Site6	-	-	-
7	Site7	-	-	-
8	Site8	-	-	-
9	Site9	-	-	-
Etc.	Etc.	-	-	-

Internal “Site Inspectors” (SIs)

- Individuals assigned a set of Sites for weekly review and for monitoring of “Data tickets” submitted through Help Desk
- SOP developed and continues to be refined by internal staff focus on key operational QA for weekly reviews



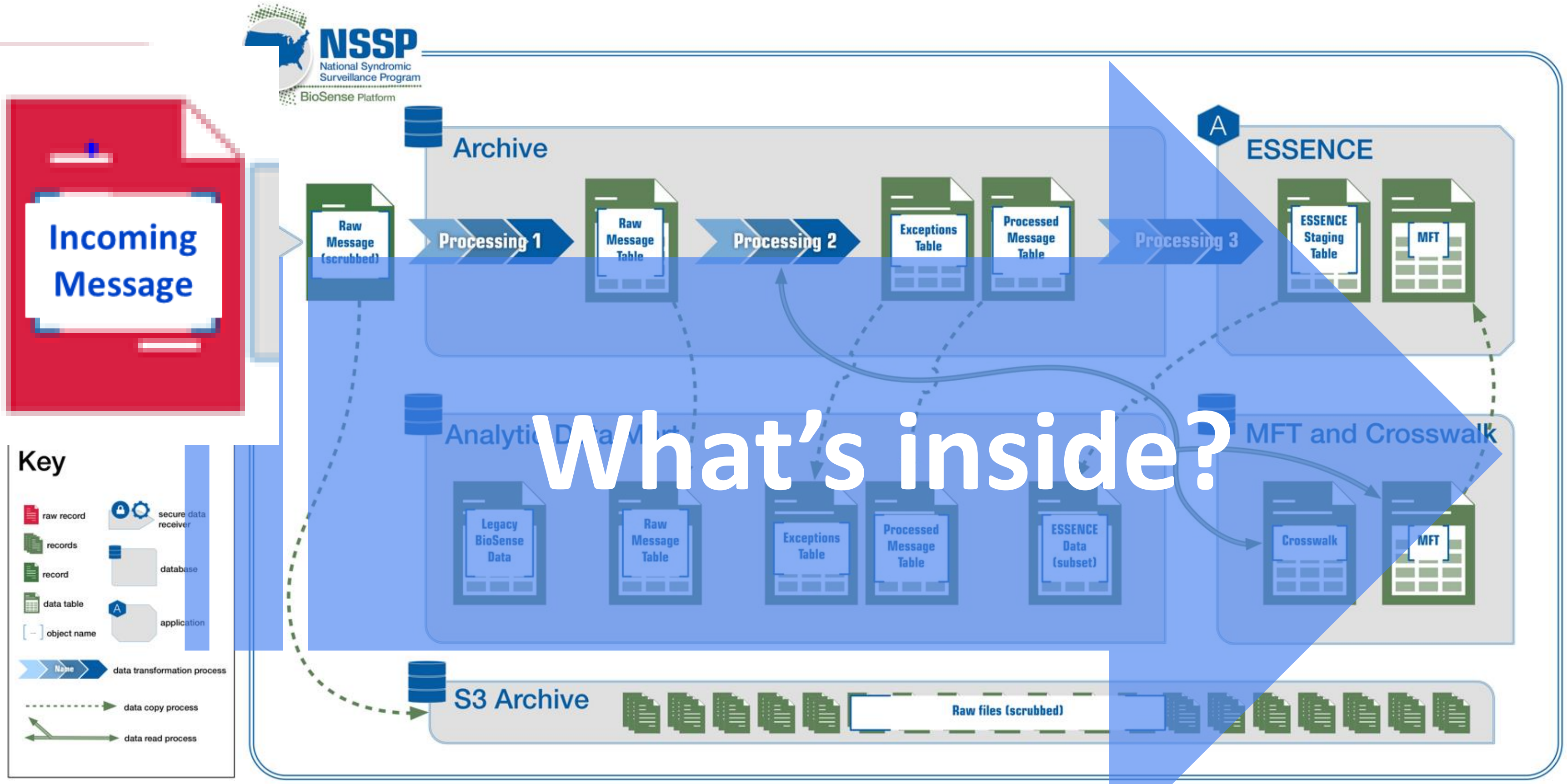
	Primary		Secondary	
Assignee	Feeds	Facilities	Feeds	Facilities
Inspector 1	50	449	39	1283
Inspector 2	87	1000	35	793
Inspector 3	13	698	78	714
Inspector 4	27	1330	25	687
Total	**177**	**3477**	**177**	**3477**

Primary and a Secondary SIs assigned among 60+ sites

Overview – Data Content

- High-level Review of Data Flow
- Foundational Data Quality (DQ)
- Deeper Dive DQ Review of Data Content
- Feedback from the Community
- Next Steps

Deeper Dive – Data Content



Incoming Message

What's inside?

Last modified: January 27, 2016

NOTE: CONCEPTUAL DIAGRAM. FOR PLANNING & DISCUSSION USE ONLY.

Data Quality Reports: Starting Point

- Beta process established to assist with internal QA of Staging Data *(during transition)*
- Reports developed for
  - Timeliness
  - Completeness
  - Validity
- Transitioned reports to run against Production Data *(post transition)* to assist with routine operational QA

Data Quality Reports: Intent

- Standardize reports across sites for internal operational QA
- Identify potential processing issues and/or incoming data issues – investigate further to “get to the root of the problem”
- Support sites that lack sufficient QA resources
- Work with the community to refine reports
- (Potentially) provide supplementary information to Grantees that will assist in generating performance measures

Reports do not supplant QA work being done by sites that have well-established QA processes

Data Quality Report: Releases

- "Soft release" of Production Data Reports to Sites (Fall 2016)
  - Emailed to site administrators
  - Invited to provide overview of reports during community webinars
  - Solicited and collected helpful feedback from the community
- Prospective monthly release of beta reports – Production Data
  - Secure File Transfer Protocol (SFTP) pickup area (January 2017)
  - Access & Management Center or other dashboards (future)
- Onboarding "Data Validation" (same code-based process generates reports that support onboarding data validation)

Reports provide data overall, by feed, or by feed and facility

Data Quality Reports

- Timeliness
- Completeness
- Validity

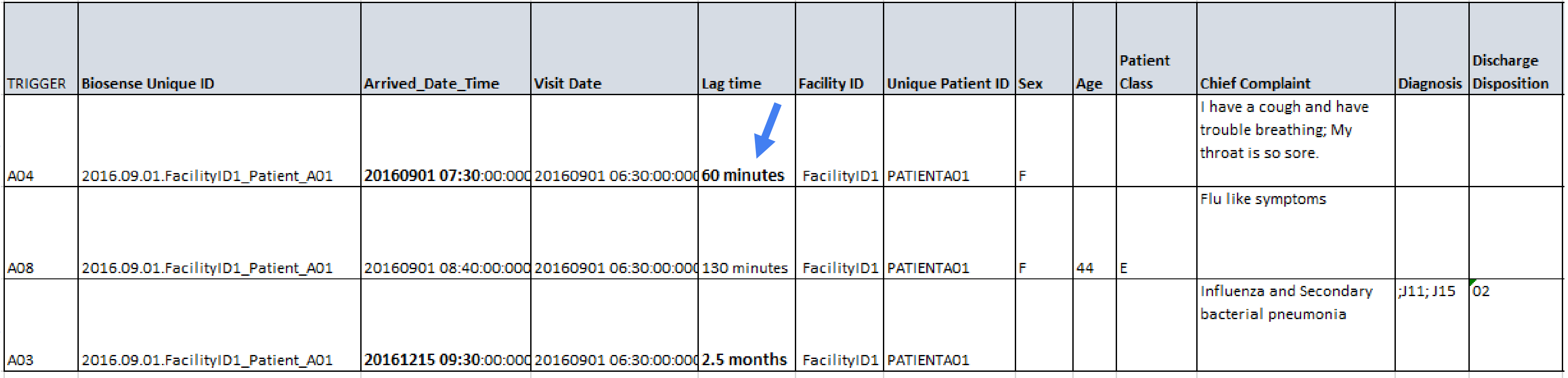
Timeliness

How long does it take the data to arrive on the platform?

- Lag time is measured from “date/time of the visit” to “date/time the first message arrived” on the BioSense Platform
- Subsequent messages for same visit are NOT considered to avoid skewing the results
- Reports include graphs and tables
- Metrics are for 24 hours and 48 hours

Example: Importance of Using First Arrival Date

- Patient visits facility on 09/01/2016, 6:30 am
- First message arrives on platform 60 minutes later at 7:30am
- Last message, with a diagnosis update, arrives about 2 ½ months later
- Although 3 physical messages were sent over time, this counts as **1 visit** with a lag time of **60 minutes**



V
I
S
I
T

TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Lag time	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A01	**20160901 07:30**:00:000	20160901 06:30:00:000	**60 minutes**	FacilityID1	PATIENTA01	F			I have a cough and have trouble breathing; My throat is so sore.		
A08	2016.09.01.FacilityID1_Patient_A01	20160901 08:40:00:000	20160901 06:30:00:000	130 minutes	FacilityID1	PATIENTA01	F	44	E	Flu like symptoms		
A03	2016.09.01.FacilityID1_Patient_A01	**20161215 09:30**:00:000	20160901 06:30:00:000	**2.5 months**	FacilityID1	PATIENTA01				Influenza and Secondary bacterial pneumonia	;J11; J15	02

Timeliness: Report Set

- Graphs include
  - Visit counts
  - Median number of days from visit to arrival over time
- Summary Tables include Timeliness Performance Categories
    - 0–<30% of visits arriving within 24 hours; within 48 hours
    - 30–<80% of visits
    - >80% of visits
- Detail Tables include
  - Timeliness Performance Categories
  - Mean/Median number of lag days
  - Lag days associated with >80% of visits

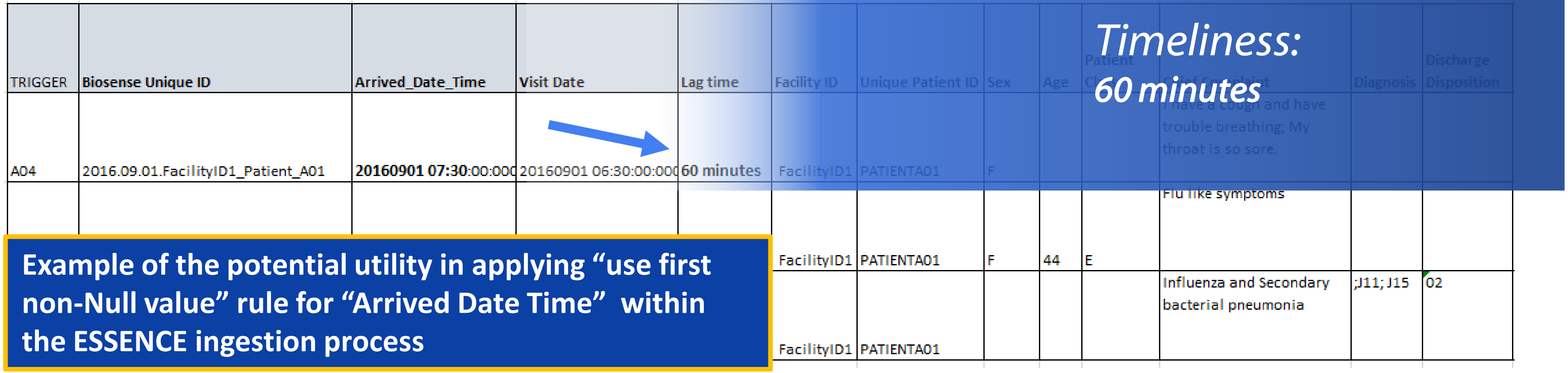
Difference in Timeliness Reports

ARCHIVE Data (DQ reports) and ESSENCE (DQ dashboard)

- DQ Reports using ARCHIVE data
  - Calculation is based on the difference between the visit date/time and the date/time that very first message arrived on the platform
- DQ Dashboard in ESSENCE
  - Data ingestion process is based on the most recently received message for the visit (with some exceptions)
  - Calculation is therefore based on the difference between the visit date/time and the most recent message date/time associated with the set of messages for that visit

Example: Difference in Timeliness
ARCHIVE Data DQ Reports and ESSENCE

DQ Reports

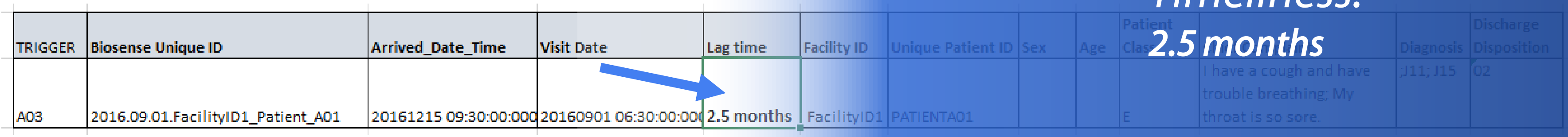


TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Lag time	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A01	20160901 07:30:00:000	20160901 06:30:00:000	60 minutes	FacilityID1	PATIENTA01	F			I have a cough and have trouble breathing; My throat is so sore.		
					FacilityID1	PATIENTA01	F	44	E	Flu like symptoms		
					FacilityID1	PATIENTA01				Influenza and Secondary bacterial pneumonia	;J11; J15	02

Timeliness:
60 minutes

Example of the potential utility in applying "use first non-Null value" rule for "Arrived Date Time" within the ESSENCE ingestion process

ESSENCE



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Lag time	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A03	2016.09.01.FacilityID1_Patient_A01	20161215 09:30:00:000	20160901 06:30:00:000	2.5 months	FacilityID1	PATIENTA01			E	I have a cough and have trouble breathing; My throat is so sore.	;J11; J15	02

Timeliness:
2.5 months

Data Quality Reports - Completeness

- Timeliness
- Completeness
- Validity

Completeness

Are data populated?

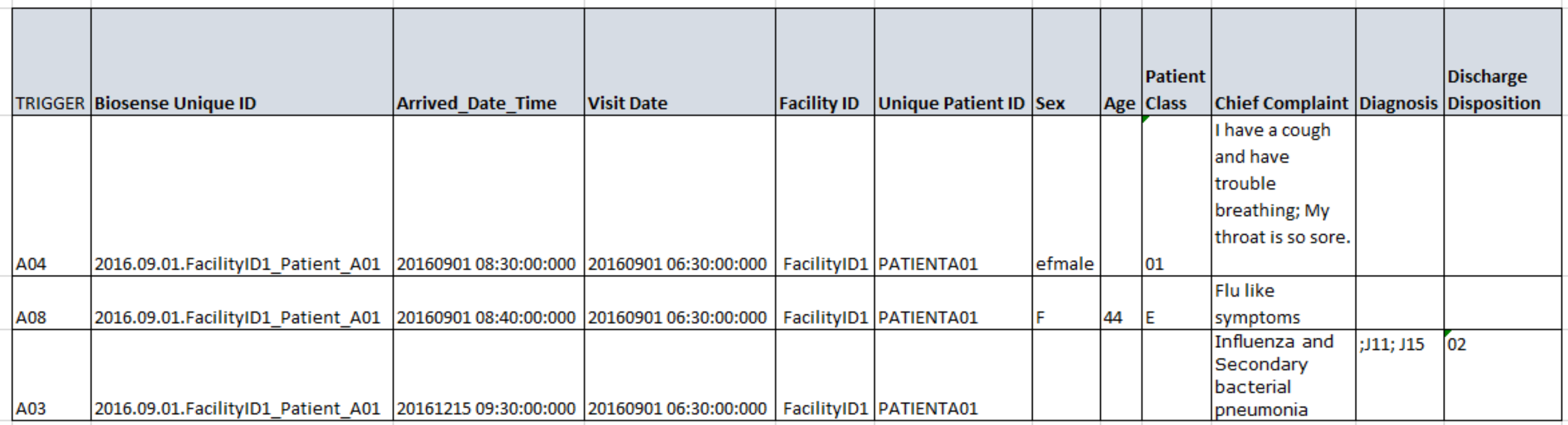
Of all the opportunities the facility had to send data for unique patient visit, for a particular data element, was it ever sent for that visit?

- Consider all records that are associated with a unique patient visit (assesses Incoming data and not the downstream process)
- Determine if a data element for a unique patient visit is complete based on whether any of the records (for the visit) carried data for that data element
- Mark as complete vs. non-complete based on what was found across records
- Calculate percent complete (for each data element) based on a visit-level denominator

Example of Visit Data: Visit level completeness

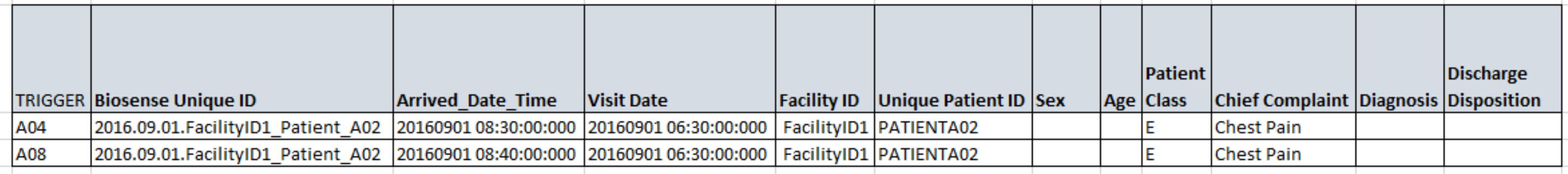
- Three records (messages) sent for a unique patient visit (Visit #1)
- Two records (messages) sent for a different unique patient visit (Visit #2)
- Some but not all of the records have data in various data elements

Visit #1



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A01	20160901 08:30:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA01	efmale		01	I have a cough and have trouble breathing; My throat is so sore.		
A08	2016.09.01.FacilityID1_Patient_A01	20160901 08:40:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA01	F	44	E	Flu like symptoms		
A03	2016.09.01.FacilityID1_Patient_A01	20161215 09:30:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA01				Influenza and Secondary bacterial pneumonia	;J11; J15	02

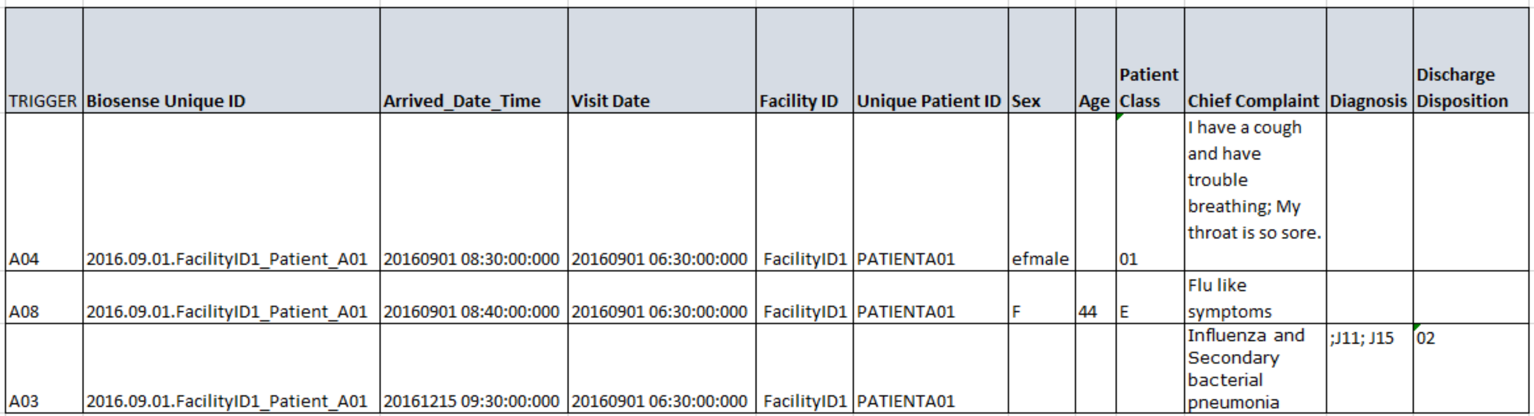
Visit #2



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A02	20160901 08:30:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA02			E	Chest Pain		
A08	2016.09.01.FacilityID1_Patient_A02	20160901 08:40:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA02			E	Chest Pain		

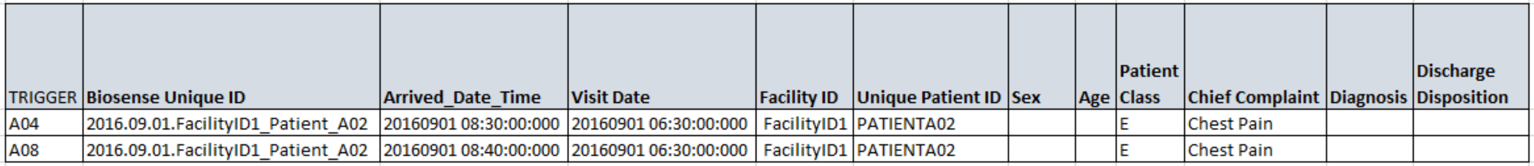
Example: Visit level completeness

Visit #1



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A01	20160901 08:30:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA01	efmale		01	I have a cough and have trouble breathing; My throat is so sore.		
A08	2016.09.01.FacilityID1_Patient_A01	20160901 08:40:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA01	F	44	E	Flu like symptoms		
A03	2016.09.01.FacilityID1_Patient_A01	20161215 09:30:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA01				Influenza and Secondary bacterial pneumonia	;J11; J15	02

Visit #2



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A02	20160901 08:30:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA02			E	Chest Pain		
A08	2016.09.01.FacilityID1_Patient_A02	20160901 08:40:00:000	20160901 06:30:00:000	FacilityID1	PATIENTA02			E	Chest Pain		

Total records:	*5*	
Total visits:	*2 (the denominator)*	
%Complete:	*Sex (50%)*	*1 of the 2 visits have data*
	Age (50%)	*1 of the 2 visits have data*
	CC (100%)	*both of the visits have data*

Columns to assist with "drill down"

- "Use Group": Categories of data elements including
  - Chief Complaint/Diagnosis
  - Demographics
  - Facility
  - Visit Information
- "Required": Usage categories including
  - R (Required)
  - RE (Required buy may be initially empty)
  - CR (Calculated by NSSP data flow, dependent on one or more "R" data elements)
  - CRE (Calculated by NSSP data flow, dependent on one or more "RE" data elements)
  - O (optional)

R, RE, CR, RE elements are highlighted if percent complete <90%

("By Trigger" reports slated for the future to support the variation in Required fields across trigger types)

- "HL7": HL-7 segments

====Same drill down columns available in Validity Reports====

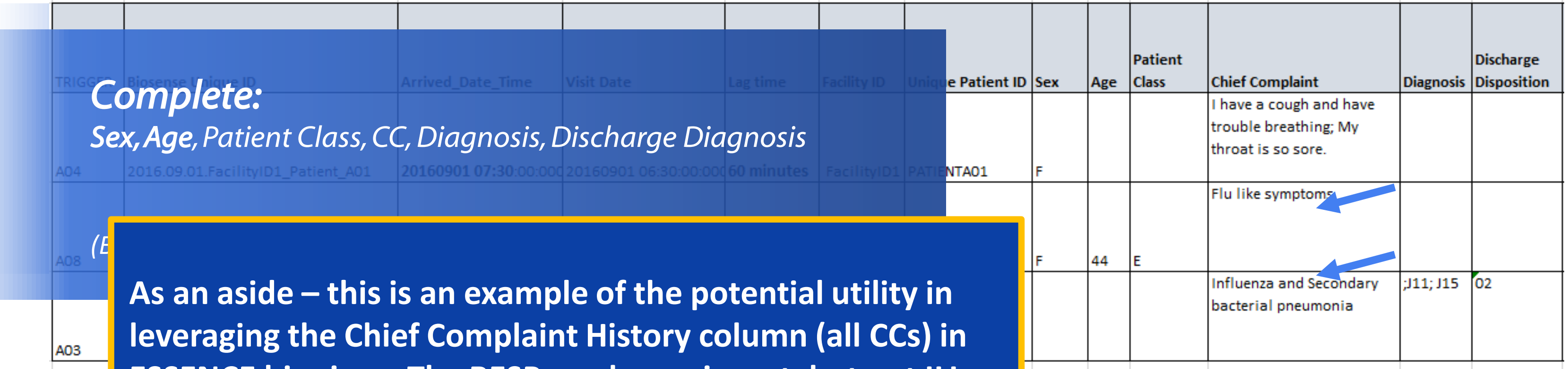
Difference in Completeness ARCHIVE Data and ESSENCE

- Data received in the most recent message is used to ingest into ESSENCE
- Exceptions include
  - Patient Class (last non-NULL)
  - Chief Complaint (first non-NULL)
  - Diagnosis (last non-NULL)
  - Discharge Disposition (last non-NULL)

Example: Difference in Completeness

ARCHIVE Data DQ Reports and ESSENCE

Data used in DQ Reports for Visit #1



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Lag time	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition
A04	2016.09.01.FacilityID1_Patient_A01	20160901 07:30:00:000	20160901 06:30:00:000	60 minutes	FacilityID1	PATIENTA01	F			I have a cough and have trouble breathing; My throat is so sore.		
A08							F	44	E	Flu like symptoms		
A03										Influenza and Secondary bacterial pneumonia	;J11; J15	02

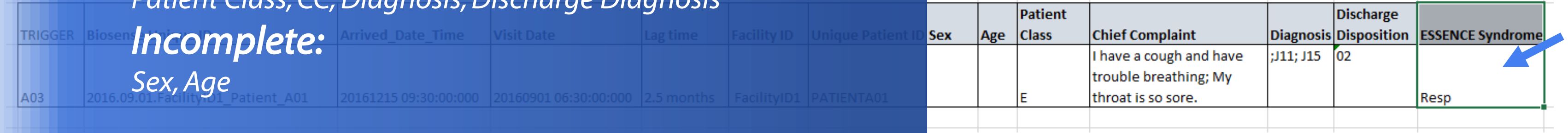
Complete:

Sex, Age, Patient Class, CC, Diagnosis, Discharge Diagnosis

(B

As an aside – this is an example of the potential utility in leveraging the Chief Complaint History column (all CCs) in ESSENCE binning; The RESP syndrome is met, but not ILI .

Data in ESSENCE for Visit #1



TRIGGER	Biosense Unique ID	Arrived_Date_Time	Visit Date	Lag time	Facility ID	Unique Patient ID	Sex	Age	Patient Class	Chief Complaint	Diagnosis	Discharge Disposition	ESSENCE Syndrome
A03	2016.09.01.FacilityID1_Patient_A01	20161215 09:30:00:000	20160901 06:30:00:000	2.5 months	FacilityID1	PATIENTA01			E	I have a cough and have trouble breathing; My throat is so sore.	;J11; J15	02	Resp

Complete:

Patient Class, CC, Diagnosis, Discharge Diagnosis

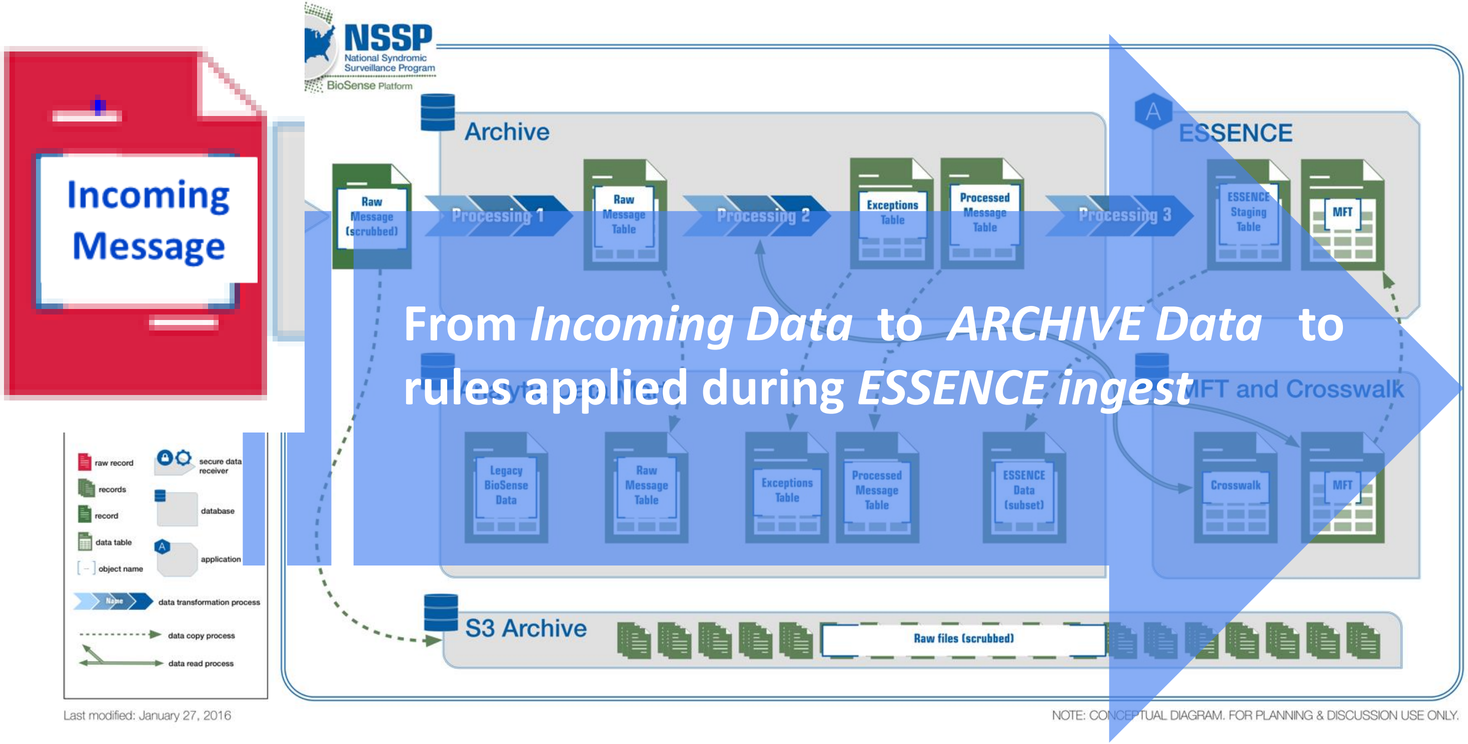
Incomplete:

Sex, Age

(Based on business rules applied in ESSENCE ingestion)

DQ for both Incoming data and ESSENCE

Serves different but equally important purposes



From *Incoming Data* to *ARCHIVE Data* to rules applied during *ESSENCE ingest*

Last modified: January 27, 2016

NOTE: CONCEPTUAL DIAGRAM. FOR PLANNING & DISCUSSION USE ONLY.

Completeness Reports: Filtered and Excepted Data

Reports include information on data that did not advance to Processed data (and therefore did not advance to ESSENCE)

- Filtered: does not meet minimum criteria of
  - ADT type message
  - Message DateTime reported
  - Sending Facility reported
- Exceptions: have one or more of the following exceptions
  - Invalid Patient ID (<3 characters or missing)
  - Invalid or missing Visit Date
  - Facility ID not registered in the MFT/Crosswalk
  - Visit Date in the future

Reports include total count and percent of filtered and excepted records; Reports include breakout of reason for triaging to filtered and excepted tables

Data Quality Reports - Validity

- Timeliness
- Completeness
- Validity

Validity

Are pertinent data elements adhering to standards?

- Targets data elements of interest that have an associated vocabulary (e.g., Administrative Sex)
- Calculates conformance at
  - "record level" (# and percent of records that conform)
  - "visit level" * (# of visits that conform)
    - Mirrors the collapsing rules used in ESSENCE ingestion to yield 1 record per each visit
- Categorizes "missing data" as non-conforming

******Facilitate assessment of incoming data as well as the data as it would appear in ESSENCE***

Validity

Other data elements

- Includes other important data elements which may not have an associated standard
- For example:
  - Age: Flag outliers
  - Initial Temperature: Flag outliers
  - Chief Complaint (CC):
    - Report out top 20 Chief Complaint Values
    - Categorize specific values as non-conforming category “CC Unk Group” *(unknown, n/a, na, unk, ed visit, ed, er, see tsheet)*
    - Categorize CC with length <= 2 as non-conforming category “CC Length LE2”

Overview Contd….

- High-level Review of Data Flow
- Foundational Data Quality (DQ)
- Deeper Dive DQ Review of Data Content
- Feedback from the Community
- Next Steps

Feedback

Slice and Dice Reports based on

- Trigger Events
  - Record level for A01, A04, A03
  - Visit level for A08
- Patient Class History Combinations
  - Emergency Visits Only
  - Emergency followed by Inpatient Admit
- Vendor

Feedback Contd.

Slice and dice based on a date range of interest for

- Arrival Date
- Message Date
- Visit Date

Feedback Contd..

- Report on Patient Age ranges “not found” in the data
- Add other “unknown” Chief Complaint checks (e.g., ?, x, XX)
- Validate diagnosis codes

Feedback Contd..

- Expand list of “units” values deemed as conforming (e.g., Temperature; Height/Weight)
- Consider unit of measure when assessing “the measure” itself:
  - Reported Age, Calculated Age
  - Temperature
  - Height, Weight
  - Blood Pressure

Overview Contd…..

- High-level Review of Data Flow
- Foundational Data Quality (DQ)
- Deeper Dive DQ Review of Data Content
- Feedback from the Community
- Next Steps

Next Steps

Design, Develop, Implement “star schema” DQ database

- Adds flexibility in “slicing and dicing”

Develop “views” into database to

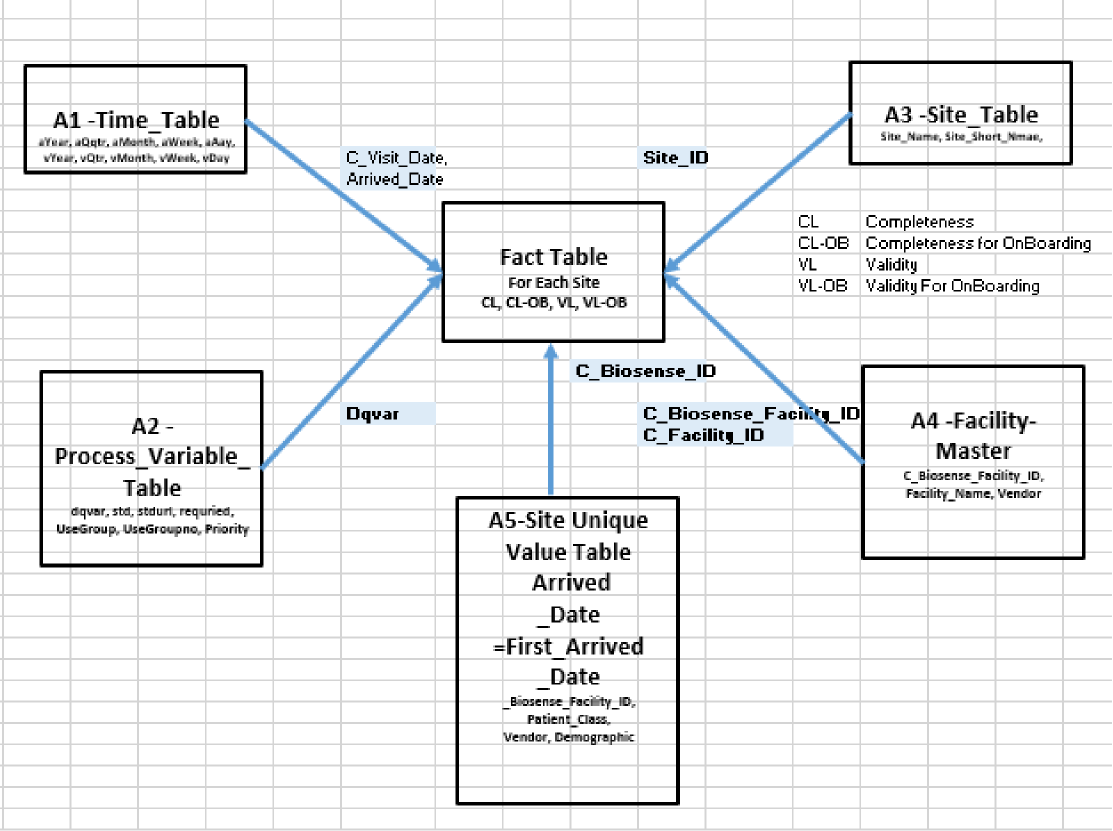
- Provide users data to query on their own, complementing end-user reports

Consider future “posting” of reports through

- Access Management Center (AMC)
- Other dashboards

Continue to work with the community as we build requirements for next phase of DQ data and reports!

Next Steps: Draft Design of “star schema”



A1 -Time_Table
aYear, aQqtr, aMonth, aWeek, aAay,
vYear, vQtr, vMonth, vWeek, vDay

C_Visit_Date,
Arrived_Date

Site_ID

A3 -Site_Table
Site_Name, Site_Short_Nmae,

Fact Table
For Each Site
CL, CL-OB, VL, VL-OB

CL	Completeness
CL-OB	Completeness for OnBoarding
VL	Validity
VL-OB	Validity For OnBoarding

C_Biosense_ID

Dqvar

C_Biosense_Facility_ID
C_Facility_ID

A2 -
Process_Variable_
Table
dqvar, std, stdurl, requried,
UseGroup, UseGroupno, Priority

A4 -Facility-
Master
C_Biosense_Facility_ID,
Facility_Name, Vendor

A5-Site Unique
Value Table
Arrived
_Date
=First_Arrived
_Date
_Biosense_Facility_ID,
Patient_Class,
Vendor, Demographic

Thank you.

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.